Mortality descriptive

Esther Chung

March 6, 2018

Goal: Identify and explore mortality variables in the GHAP nutritional intervention trials.

# According to kikm\_variables, the following nutritional intervention studies have AGEDTH as mortality variables   
 # DIVIDS, JIVITA3 and 4, iDOSE, iDYAD, ZVITAMBO, TanzaniaChild2, PROVIDE  
  
# DIVIDS  
 table(dvds$DEAD)

##   
## 1   
## 92

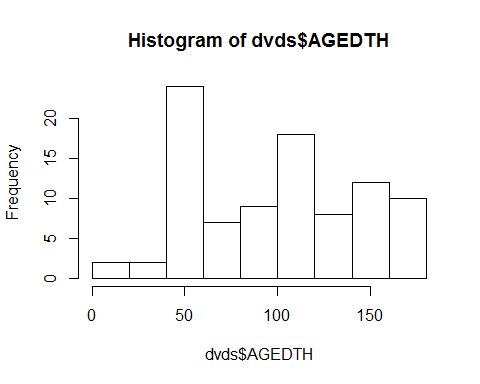
dvds$AGEDTH <- as.numeric(dvds$AGEDTH)  
 table(!is.na(dvds$AGEDTH)) # 92 deaths

##   
## FALSE TRUE   
## 20496 92

summary(dvds$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 14.0 60.0 109.0 100.6 128.0 180.0 20496

hist(dvds$AGEDTH)



# Also has cause of death(CAUSEDTH0 and contributing cause of death(CAUSDTHC)  
  
 # Check if age of death is recorded multiple times for each child who died  
 dvds %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 16 x 2  
## SUBJID N  
## <int> <int>  
## 1 101 1  
## 2 124 10  
## 3 143 4  
## 4 259 7  
## 5 260 12  
## 6 331 8  
## 7 401 9  
## 8 487 3  
## 9 529 9  
## 10 714 1  
## 11 801 7  
## 12 1098 1  
## 13 1208 8  
## 14 1418 1  
## 15 1467 1  
## 16 1666 10

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 dvds <- dvds %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 dvds <- dvds %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child if AGEDTH is not missing and mark the last observation per child  
 dvds <- dvds %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(dvds$deathErrorFlag[!is.na(dvds$AGEDTH)]) # 8

##   
## FALSE TRUE   
## 84 8

#ISSUE - 8 children have measurements after their recorded age of death  
   
 # subset data to children with death error flags  
 df\_error <- dvds %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 8 x 6  
## # Groups: SUBJID [8]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <int> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 259 T 60.0 182 182 T   
## 2 260 T 142 182 182 T   
## 3 331 T 109 182 182 T   
## 4 401 T 99.0 182 182 T   
## 5 529 T 47.0 182 182 T   
## 6 801 T 51.0 182 182 T   
## 7 1208 T 128 182 182 T   
## 8 1666 T 114 182 182 T

#make indicator for any death   
 dvds <- dvds %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
 #subset data set to 1) children who died AND 2) do not have death error flags  
 df\_death <- dvds %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, deathErrorFlag))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [4]  
## SUBJID AGEDAYS WHZ AGEDTH died deathErrorFlag  
## <int> <int> <dbl> <dbl> <lgl> <lgl>   
## 1 101 1 - 3.48 22.0 T F   
## 2 124 1 - 2.46 180 T F   
## 3 124 72 NA 180 T F   
## 4 124 74 - 5.60 180 T F   
## 5 124 100 NA 180 T F   
## 6 124 105 - 5.35 180 T F   
## 7 124 126 - 5.85 180 T F   
## 8 124 128 NA 180 T F   
## 9 124 156 NA 180 T F   
## 10 124 158 - 3.35 180 T F   
## 11 124 179 - 5.21 180 T F   
## 12 143 1 - 4.10 78.0 T F   
## 13 143 43 NA 78.0 T F   
## 14 143 71 NA 78.0 T F   
## 15 143 75 - 5.20 78.0 T F   
## 16 259 1 - 1.47 60.0 T F   
## 17 259 44 - 3.32 60.0 T F   
## 18 259 70 NA 60.0 T F   
## 19 259 98 NA 60.0 T F   
## 20 259 126 NA 60.0 T F

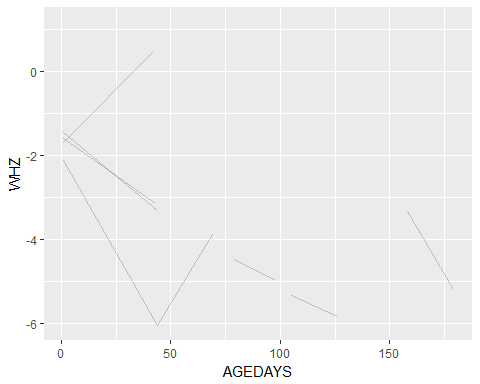
#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

## [1] -3.48 -2.46 NA -5.60 NA -5.35 -5.85 NA NA -3.35 -5.21  
## [12] -4.10 NA NA -5.20 -1.47 -3.32 NA NA NA NA -1.35  
## [23] NA NA -4.27 NA -4.48 -4.98 NA NA -4.69 NA NA  
## [34] -2.09 NA -3.31 NA NA NA NA 0.66 NA -0.65 NA  
## [45] NA NA NA NA -0.68 NA NA NA NA 1.16 NA  
## [56] NA NA NA -0.10 -1.61 -3.14 NA NA NA NA NA  
## [67] -2.13 -6.06 -3.90 NA NA NA NA NA -0.79 -1.70 0.48  
## [78] NA -1.58 NA NA -1.36 NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH))

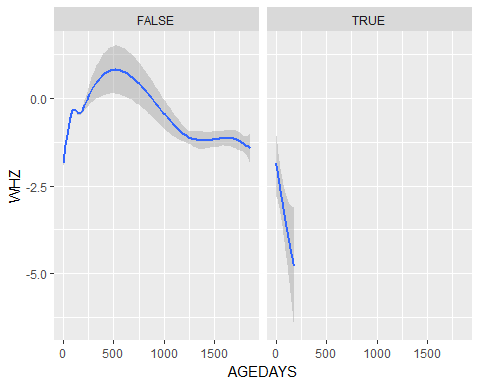
##   
## FALSE TRUE   
## 51 33

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)



# warning removed 33 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(dvds) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'



# warning removed 9,733 rows containing non-finite values

# JiVitA-3  
 jvit3$AGEDTH <- as.numeric(jvit3$AGEDTH)  
 table(!is.na(jvit3$AGEDTH)) #2515 deaths

##   
## FALSE TRUE   
## 130906 2515

summary(jvit3$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1.0 20.0 202.0 270.7 447.5 1226.0 130906

# Check if age of death is recorded multiple times for each child who died  
 jvit3 %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 1,031 x 2  
## SUBJID N  
## <int> <int>  
## 1 21 1  
## 2 26 1  
## 3 48 1  
## 4 59 1  
## 5 61 1  
## 6 73 1  
## 7 232 1  
## 8 245 1  
## 9 250 1  
## 10 251 1  
## 11 271 1  
## 12 275 1  
## 13 291 2  
## 14 296 1  
## 15 365 1  
## 16 369 1  
## 17 389 3  
## 18 426 1  
## 19 433 2  
## 20 471 2  
## 21 504 2  
## 22 506 1  
## 23 512 1  
## 24 537 1  
## 25 586 1  
## 26 605 3  
## 27 608 2  
## 28 643 1  
## 29 682 1  
## 30 686 1  
## 31 695 4  
## 32 739 3  
## 33 759 1  
## 34 791 4  
## 35 792 3  
## 36 815 7  
## 37 838 1  
## 38 899 1  
## 39 912 1  
## 40 923 1  
## # ... with 991 more rows

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(jvit3$deathErrorFlag[!is.na(jvit3$AGEDTH)]) # 81

##   
## FALSE TRUE   
## 2434 81

#ISSUE- 81 children have measurements after their recorded age of death  
   
 # subset data to children with death error flags  
 df\_error <- jvit3 %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <int> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 815 T 62.0 194 194 T   
## 2 948 T 4.00 201 201 T   
## 3 1735 T 1.00 202 202 T   
## 4 1773 T 1.00 183 183 T   
## 5 1827 T 4.00 183 183 T   
## 6 2232 T 78.0 188 188 T   
## 7 2377 T 10.0 181 181 T   
## 8 2683 T 1.00 188 188 T   
## 9 2932 T 3.00 184 184 T   
## 10 3106 T 17.0 196 196 T   
## 11 3186 T 1.00 195 195 T   
## 12 4305 T 1.00 180 180 T   
## 13 5571 T 3.00 183 183 T   
## 14 6625 T 1.00 187 187 T   
## 15 6662 T 12.0 188 188 T   
## 16 6739 T 127 183 183 T   
## 17 6744 T 2.00 182 182 T   
## 18 7679 T 4.00 183 183 T   
## 19 7845 T 1.00 203 203 T   
## 20 8266 T 1.00 181 181 T

#make indicator for any death  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
#subset data set to 1) children who died AND 2) do not have death error flags  
 df\_death <- jvit3 %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, deathErrorFlag))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [17]  
## SUBJID AGEDAYS WHZ AGEDTH died deathErrorFlag  
## <int> <int> <dbl> <dbl> <lgl> <lgl>   
## 1 21 1 NA 2.00 T F   
## 2 26 1 NA 22.0 T F   
## 3 48 1 NA 10.0 T F   
## 4 59 1 NA 3.00 T F   
## 5 61 1 NA 31.0 T F   
## 6 73 1 - 0.440 7.00 T F   
## 7 232 2 NA 26.0 T F   
## 8 245 1 NA 3.00 T F   
## 9 250 1 - 2.48 5.00 T F   
## 10 251 1 NA 4.00 T F   
## 11 271 1 - 1.98 36.0 T F   
## 12 275 1 NA 3.00 T F   
## 13 291 6 - 1.10 600 T F   
## 14 291 82 4.45 600 T F   
## 15 296 2 NA 3.00 T F   
## 16 365 1 NA 2.00 T F   
## 17 369 1 NA 7.00 T F   
## 18 389 1 NA 136 T F   
## 19 389 31 - 0.220 136 T F   
## 20 389 98 1.63 136 T F

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

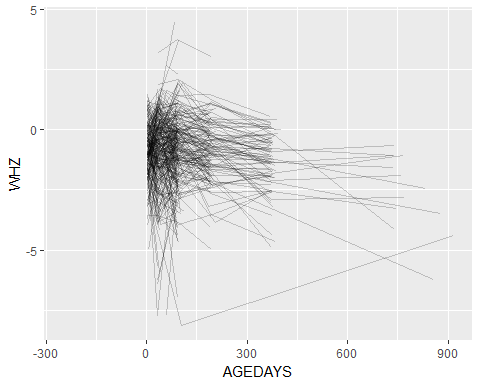
## [1] NA NA NA NA NA -0.44 NA NA -2.48 NA -1.98  
## [12] NA -1.10 4.45 NA NA NA NA -0.22 1.63 NA -1.08  
## [23] -2.00 -0.41 -0.98 -0.15 0.07 NA NA NA -1.91 -1.61 -1.61  
## [34] -3.44 -0.99 -1.20 NA NA NA -1.73 -0.13 -0.04 0.19 NA  
## [45] -1.32 NA NA -0.96 -2.59 -1.60 -2.08 NA -1.63 -2.71 NA  
## [56] NA -0.64 NA -1.40 NA NA NA -1.81 NA -1.15 -1.49  
## [67] -2.29 -1.78 NA NA NA NA NA -0.75 0.13 -0.89 -1.14  
## [78] -0.70 -0.24 -1.12 NA 1.37 -0.47 -3.18 -2.54 -1.02 -0.82 NA  
## [89] NA -0.83 -1.24 -1.60 -0.82 -0.29 NA NA 0.27 -0.11 -0.11  
## [100] -2.87 -1.85 NA -2.75 -2.16 -2.02 NA NA -5.29 NA NA  
## [111] -4.01 -0.19 NA NA -0.27 -1.28 NA -0.73 NA -0.88 -1.25  
## [122] NA NA NA NA -1.39 -1.33 -1.51 -1.07 NA NA NA  
## [133] -1.60 -3.07 NA NA -2.65 -2.28 NA NA NA NA -2.45  
## [144] -2.50 -2.80 NA NA NA NA NA 0.97 0.24 NA NA  
## [155] NA -1.62 NA -2.29 -1.88 NA NA NA NA NA NA  
## [166] -2.27 NA NA -1.98 -0.30 NA NA -0.78 -0.78 -0.57 NA  
## [177] NA 0.48 NA NA -0.56 -1.96 NA NA NA NA -3.44  
## [188] -2.79 NA -2.44 NA NA NA NA NA -1.40 NA -0.90  
## [199] -1.21 NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 943 missing WHZ

##   
## FALSE TRUE   
## 943 1491

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

## Warning: Removed 307 rows containing missing values (geom\_path).



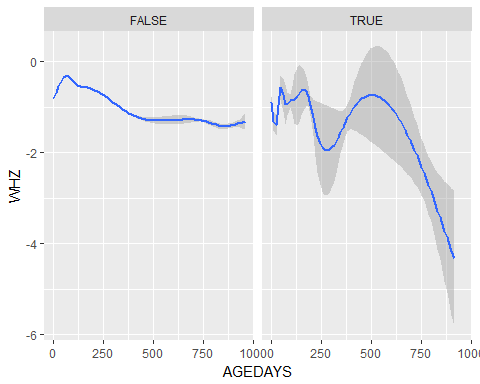
# warning indicates that 307 rows had missing values   
   
 # check how obs are missing WHZ   
 table(!is.na(df\_death$WHZ)) # missing 943 WHZ measurements

##   
## FALSE TRUE   
## 943 1491

#Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(jvit3) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 16246 rows containing non-finite values (stat\_smooth).



# warning indicates that 16,246 rows contained non-finite values

# JiVitA-4  
 table(jvit4$DEAD) # indicator; 468 deaths

##   
## 0 1   
## 83682 468

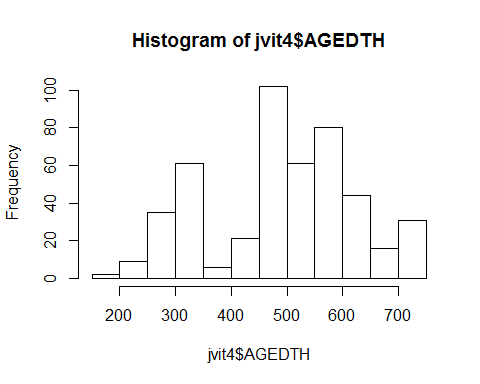
jvit4$AGEDTH <- as.numeric(jvit4$AGEDTH)  
 table(!is.na(jvit4$AGEDTH)) #468 deaths

##   
## FALSE TRUE   
## 83682 468

summary(jvit4$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 187.0 405.0 500.0 492.6 582.5 735.0 83682

hist(jvit4$AGEDTH)



# Check if age of death is recorded multiple times for each child who died  
 jvit4 %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 50 x 2  
## SUBJID N  
## <int> <int>  
## 1 23 3  
## 2 263 12  
## 3 323 3  
## 4 704 7  
## 5 823 16  
## 6 849 16  
## 7 861 6  
## 8 902 11  
## 9 922 4  
## 10 1021 11  
## 11 1282 17  
## 12 1474 12  
## 13 1591 6  
## 14 1718 10  
## 15 1768 6  
## 16 1835 5  
## 17 1843 13  
## 18 1864 6  
## 19 1872 12  
## 20 1934 12  
## 21 2220 12  
## 22 2303 14  
## 23 2363 6  
## 24 2395 13  
## 25 2501 6  
## 26 2619 4  
## 27 2630 17  
## 28 2774 13  
## 29 2862 2  
## 30 2924 11  
## 31 2946 4  
## 32 3069 7  
## 33 3327 9  
## 34 3349 11  
## 35 3421 16  
## 36 3921 8  
## 37 4020 7  
## 38 4245 6  
## 39 4377 12  
## 40 4483 12  
## # ... with 10 more rows

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(jvit4$deathErrorFlag[!is.na(jvit4$AGEDTH)]) # 0! :D

##   
## FALSE   
## 468

#make indicator for any death  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
#subset data set to 1) children who died AND 2) do not have death error flags  
 df\_death <- jvit4 %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died))  
   
 head(df\_death, 20)

## # A tibble: 20 x 5  
## # Groups: SUBJID [4]  
## SUBJID AGEDAYS WHZ AGEDTH died   
## <int> <int> <dbl> <dbl> <lgl>  
## 1 23 177 - 0.660 287 T   
## 2 23 215 - 1.95 287 T   
## 3 23 277 - 3.49 287 T   
## 4 263 8 - 3.03 514 T   
## 5 263 126 0.720 514 T   
## 6 263 187 - 1.38 514 T   
## 7 263 243 NA 514 T   
## 8 263 273 NA 514 T   
## 9 263 291 - 1.65 514 T   
## 10 263 304 NA 514 T   
## 11 263 365 NA 514 T   
## 12 263 369 - 2.69 514 T   
## 13 263 425 NA 514 T   
## 14 263 460 - 2.83 514 T   
## 15 263 486 NA 514 T   
## 16 323 6 - 2.46 226 T   
## 17 323 97 - 1.35 226 T   
## 18 323 189 - 0.680 226 T   
## 19 704 4 - 1.76 346 T   
## 20 704 97 - 0.300 346 T

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

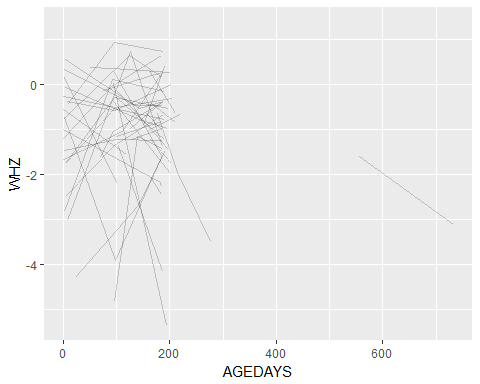
## [1] -0.66 -1.95 -3.49 -3.03 0.72 -1.38 NA NA -1.65 NA NA  
## [12] -2.69 NA -2.83 NA -2.46 -1.35 -0.68 -1.76 -0.30 -0.53 NA  
## [23] NA NA NA -0.44 0.93 0.72 NA NA NA -0.35 NA  
## [34] NA -1.00 NA -1.24 NA NA NA -1.06 0.26 -1.62 NA  
## [45] NA NA -2.67 NA NA -3.48 NA NA -2.31 NA NA  
## [56] NA -1.86 -1.47 -1.23 -1.26 NA NA -2.62 -1.67 -1.05 0.25  
## [67] NA NA NA -0.26 NA NA 0.07 NA -1.23 -1.77 NA  
## [78] NA NA -4.83 -1.16 -1.43 NA NA NA -1.49 NA NA  
## [89] -1.78 NA -0.65 -1.03 NA NA NA -1.35 NA NA -0.55  
## [100] NA -1.41 NA NA NA -1.60 -3.11 -1.17 -0.82 NA NA  
## [111] NA -0.79 NA NA -1.53 NA -1.43 NA NA -0.02 0.64  
## [122] NA NA -1.54 -0.76 0.63 0.15 NA NA NA -0.28 NA  
## [133] NA -0.52 -0.33 NA NA NA -2.26 NA -1.63 -1.04 -0.38  
## [144] NA NA -0.74 -3.91 -1.62 NA NA NA -1.87 NA NA  
## [155] NA -1.73 NA -1.70 -0.40 -0.85 NA -0.92 NA NA -2.82  
## [166] 0.10 -0.17 NA NA NA -1.10 NA -1.13 NA -1.44 NA  
## [177] -0.39 -0.59 0.00 NA NA NA -0.78 NA NA -0.41 NA  
## [188] -0.78 NA -0.99 -0.41 NA NA NA -0.40 NA NA -1.21  
## [199] NA -1.21

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 255 missing WHZ

##   
## FALSE TRUE   
## 255 213

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

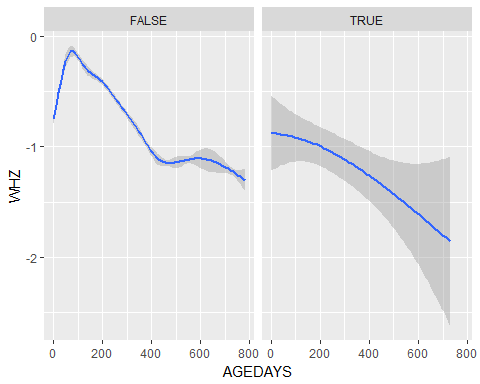
## Warning: Removed 42 rows containing missing values (geom\_path).



# warning 42 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(jvit4) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 45331 rows containing non-finite values (stat\_smooth).



# warning Removed 45331 rows containing non-finite values

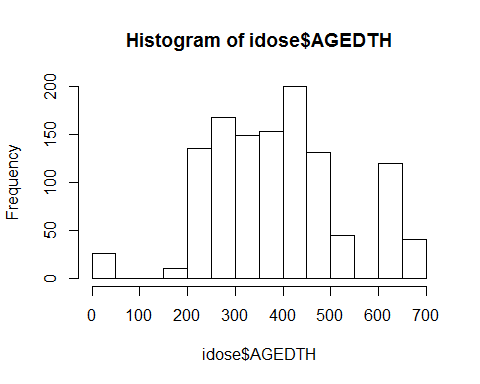
# iLiNS-DOSE  
 idose$AGEDTH <- as.numeric(idose$AGEDTH)  
 table(!is.na(idose$AGEDTH)) #1177 deaths (doesn't match DEAD)

##   
## FALSE TRUE   
## 87218 1177

summary(idose$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 7.0 285.0 385.0 386.4 459.0 667.0 87218

hist(idose$AGEDTH)



# Note: DEAD doesn't match table count using AGEDTH or the died variable   
   
 # Check if age of death is recorded multiple times for each child who died  
 idose %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 53 x 2  
## SUBJID N  
## <int> <int>  
## 1 19 59  
## 2 53 16  
## 3 77 18  
## 4 79 7  
## 5 132 14  
## 6 183 16  
## 7 197 19  
## 8 212 4  
## 9 217 15  
## 10 249 45  
## 11 379 38  
## 12 390 7  
## 13 419 9  
## 14 461 36  
## 15 488 40  
## 16 494 38  
## 17 532 48  
## 18 544 45  
## 19 557 19  
## 20 613 33  
## 21 711 61  
## 22 799 19  
## 23 801 22  
## 24 840 15  
## 25 845 9  
## 26 866 23  
## 27 880 41  
## 28 898 3  
## 29 985 26  
## 30 1002 16  
## 31 1016 29  
## 32 1051 31  
## 33 1064 40  
## 34 1074 23  
## 35 1093 4  
## 36 1134 25  
## 37 1136 8  
## 38 1204 13  
## 39 1211 19  
## 40 1215 7  
## # ... with 13 more rows

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 idose <- idose %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(idose$deathErrorFlag[!is.na(idose$AGEDTH)]) # 47

##   
## FALSE TRUE   
## 1130 47

#ISSUE - 47 children have measurements after their recorded age of death  
  
 # subset data to children with death error flags  
 df\_error <- idose %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <int> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 19 T 646 675 675 T   
## 2 53 T 323 830 830 T   
## 3 77 T 265 286 286 T   
## 4 132 T 257 326 326 T   
## 5 183 T 283 428 428 T   
## 6 197 T 285 300 300 T   
## 7 212 T 198 588 588 T   
## 8 217 T 254 262 262 T   
## 9 249 T 483 1145 1145 T   
## 10 379 T 468 498 498 T   
## 11 419 T 224 590 590 T   
## 12 461 T 409 1441 1441 T   
## 13 494 T 434 451 451 T   
## 14 544 T 417 706 706 T   
## 15 557 T 374 457 457 T   
## 16 613 T 385 399 399 T   
## 17 711 T 645 838 838 T   
## 18 799 T 315 329 329 T   
## 19 801 T 358 371 371 T   
## 20 840 T 249 263 263 T

#make indicator for any death  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- idose %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, DEAD))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [1]  
## SUBJID AGEDAYS WHZ AGEDTH died DEAD  
## <int> <int> <dbl> <dbl> <lgl> <int>  
## 1 19 198 - 1.38 646 T 1  
## 2 19 206 NA 646 T 1  
## 3 19 213 NA 646 T 1  
## 4 19 234 NA 646 T 1  
## 5 19 241 NA 646 T 1  
## 6 19 248 NA 646 T 1  
## 7 19 255 NA 646 T 1  
## 8 19 262 NA 646 T 1  
## 9 19 269 NA 646 T 1  
## 10 19 276 NA 646 T 1  
## 11 19 281 NA 646 T 1  
## 12 19 283 NA 646 T 1  
## 13 19 290 NA 646 T 1  
## 14 19 297 NA 646 T 1  
## 15 19 304 NA 646 T 1  
## 16 19 306 NA 646 T 1  
## 17 19 311 NA 646 T 1  
## 18 19 318 NA 646 T 1  
## 19 19 325 NA 646 T 1  
## 20 19 330 NA 646 T 1

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

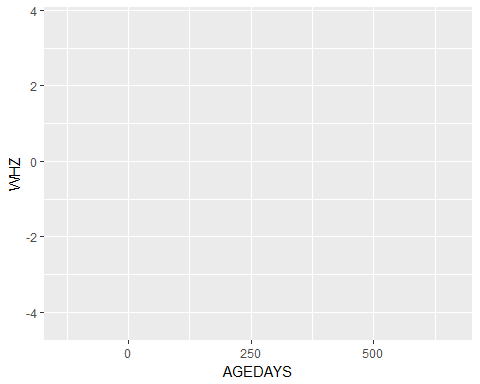
## [1] -1.38 NA NA NA NA NA NA NA NA NA NA  
## [12] NA NA NA NA NA NA NA NA NA NA NA  
## [23] NA NA NA NA NA 0.06 NA NA NA NA NA  
## [34] NA NA NA NA NA NA NA NA NA NA NA  
## [45] NA NA NA NA NA NA NA NA NA NA -2.25  
## [56] NA NA NA NA -0.85 NA NA NA NA NA NA  
## [67] NA NA NA NA NA NA NA -0.42 NA NA NA  
## [78] NA NA NA NA NA NA NA NA NA NA NA  
## [89] NA NA NA 1.05 NA NA NA NA NA NA 0.12  
## [100] NA NA NA NA NA NA NA NA NA NA NA  
## [111] -1.15 NA NA NA NA NA NA NA NA NA NA  
## [122] NA NA NA NA NA 2.30 NA NA NA NA NA  
## [133] NA NA NA NA NA NA NA NA NA NA NA  
## [144] NA -0.68 NA NA 0.91 NA NA NA NA NA NA  
## [155] NA NA NA NA NA NA NA -0.25 NA NA NA  
## [166] NA NA NA NA NA NA NA NA NA NA NA  
## [177] NA NA NA NA NA NA NA NA NA NA NA  
## [188] NA NA NA NA NA NA NA NA NA NA NA  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 1064 missing WHZ

##   
## FALSE TRUE   
## 1064 66

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

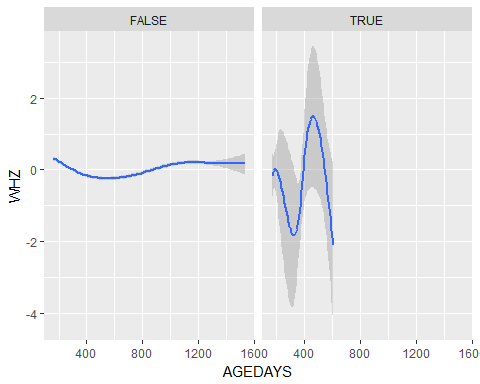
## Warning: Removed 683 rows containing missing values (geom\_path).



# Warning removed 683 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(idose) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 78964 rows containing non-finite values (stat\_smooth).



# warning removed 78964 rows containing non-finite values

# iLiNS-DYAD  
 table(idyad$DEAD) # indicator, 1679

##   
## 0 1   
## 62154 1679

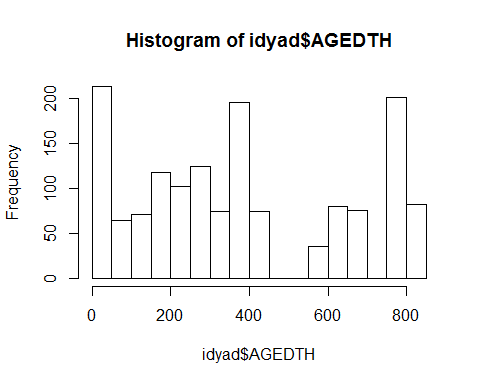
idyad$AGEDTH <- as.numeric(idyad$AGEDTH)  
 table(!is.na(idyad$AGEDTH)) #1516 deaths (doesn't match DEAD)

##   
## FALSE TRUE   
## 62317 1516

summary(idyad$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.4 162.5 336.4 372.5 601.5 824.1 62317

hist(idyad$AGEDTH)



# Note: DEAD doesn't match table count or the died variable   
   
 # Check if age of death is recorded multiple times for each child who died  
 idyad %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 59 x 2  
## SUBJID N  
## <int> <int>  
## 1 30291 34  
## 2 30331 57  
## 3 30401 9  
## 4 30542 18  
## 5 31041 88  
## 6 31061 4  
## 7 31081 10  
## 8 31281 37  
## 9 31661 5  
## 10 31911 17  
## 11 31971 41  
## 12 32081 76  
## 13 32461 31  
## 14 32661 30  
## 15 32721 36  
## 16 32781 52  
## 17 33021 10  
## 18 33091 30  
## 19 33281 39  
## 20 33381 80  
## 21 33451 19  
## 22 33462 22  
## 23 33511 23  
## 24 33611 8  
## 25 33631 9  
## 26 33791 64  
## 27 33901 82  
## 28 34081 12  
## 29 34101 13  
## 30 34111 12  
## 31 34161 7  
## 32 34291 13  
## 33 34461 7  
## 34 34581 10  
## 35 34621 5  
## 36 34921 8  
## 37 40031 37  
## 38 40321 19  
## 39 40361 12  
## 40 40461 10  
## # ... with 19 more rows

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 idyad <- idyad %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 idyad <- idyad %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child  
 idyad <- idyad %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(idyad$deathErrorFlag[!is.na(idyad$AGEDTH)]) #39

##   
## FALSE TRUE   
## 1477 39

#ISSUE - 39 children have measurements after their recorded age of death  
   
 # subset data to children with death error flags  
 df\_error <- idyad %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <int> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 30291 T 336 690 690 T   
## 2 30401 T 10.9 185 185 T   
## 3 30542 T 0.800 944 944 T   
## 4 31061 T 210 524 524 T   
## 5 31081 T 3.30 36 36.0 T   
## 6 31661 T 0.400 121 121 T   
## 7 31911 T 76.8 509 509 T   
## 8 31971 T 301 438 438 T   
## 9 32461 T 187 370 370 T   
## 10 32661 T 164 431 431 T   
## 11 32721 T 237 526 526 T   
## 12 32781 T 360 569 569 T   
## 13 33021 T 1.70 679 679 T   
## 14 33091 T 157 273 273 T   
## 15 33281 T 214 571 571 T   
## 16 33451 T 103 329 329 T   
## 17 33462 T 29.9 917 917 T   
## 18 33611 T 1.70 374 374 T   
## 19 33631 T 33.9 197 197 T   
## 20 34101 T 0.700 187 187 T

#make indicator for any death  
 idyad <- idyad %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- idyad %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, DEAD))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [1]  
## SUBJID AGEDAYS WHZ AGEDTH died DEAD  
## <int> <int> <dbl> <dbl> <lgl> <int>  
## 1 30291 -134 NA 336 T 1  
## 2 30291 -126 NA 336 T 1  
## 3 30291 -119 NA 336 T 1  
## 4 30291 -104 NA 336 T 1  
## 5 30291 - 62 NA 336 T 1  
## 6 30291 - 22 NA 336 T 1  
## 7 30291 1 NA 336 T 1  
## 8 30291 34 1.21 336 T 1  
## 9 30291 95 NA 336 T 1  
## 10 30291 110 NA 336 T 1  
## 11 30291 112 NA 336 T 1  
## 12 30291 126 NA 336 T 1  
## 13 30291 146 NA 336 T 1  
## 14 30291 152 NA 336 T 1  
## 15 30291 173 NA 336 T 1  
## 16 30291 187 - 0.760 336 T 1  
## 17 30291 188 NA 336 T 1  
## 18 30291 208 NA 336 T 1  
## 19 30291 214 NA 336 T 1  
## 20 30291 229 NA 336 T 1

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

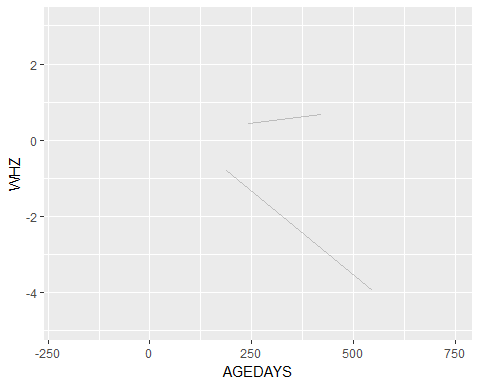
## [1] NA NA NA NA NA NA NA 1.21 NA NA NA  
## [12] NA NA NA NA -0.76 NA NA NA NA NA NA  
## [23] NA NA NA NA NA NA NA NA NA NA NA  
## [34] NA NA NA NA NA 0.11 NA NA NA NA NA  
## [45] NA NA NA NA NA NA 0.45 NA NA NA NA  
## [56] NA NA NA NA NA NA NA NA NA NA NA  
## [67] NA NA NA NA NA NA NA NA NA NA NA  
## [78] -1.84 NA NA NA NA NA NA NA NA NA NA  
## [89] NA NA NA NA NA NA NA NA NA NA NA  
## [100] NA NA NA NA NA NA NA NA NA NA NA  
## [111] NA NA NA NA NA NA NA NA NA NA NA  
## [122] NA NA NA NA -1.90 NA NA NA NA NA NA  
## [133] NA NA NA NA NA NA NA NA NA NA NA  
## [144] NA NA NA NA NA -4.70 NA NA NA NA NA  
## [155] NA NA NA NA NA NA NA NA NA NA NA  
## [166] NA NA NA NA NA NA NA -1.69 NA NA NA  
## [177] NA NA NA NA NA NA NA NA NA NA NA  
## [188] NA NA NA NA NA NA NA NA NA NA NA  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 1396 missing WHZ

##   
## FALSE TRUE   
## 1396 81

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

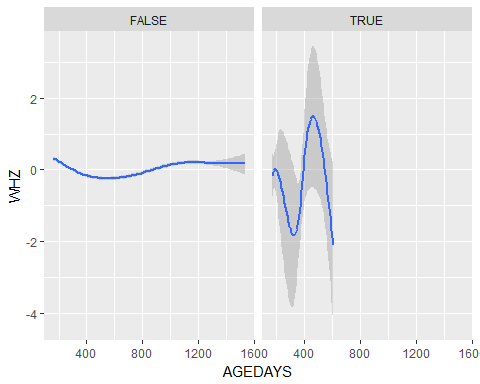
## Warning: Removed 632 rows containing missing values (geom\_path).



# warning removed 632 rows containing missing values   
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(idose) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 78964 rows containing non-finite values (stat\_smooth).



# warning Removed 78964 rows containing non-finite values

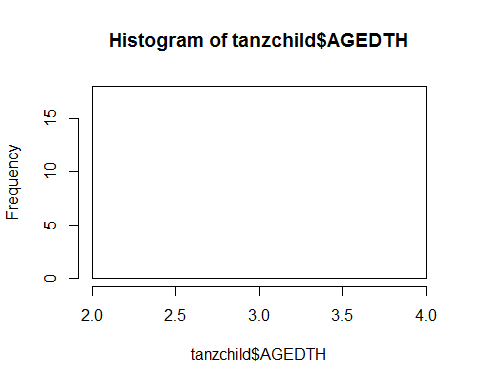
# TanzaniaChild2  
 tanzchild$AGEDTH <- as.numeric(tanzchild$AGEDTH)  
 table(!is.na(tanzchild$AGEDTH)) #18 deaths

##   
## FALSE TRUE   
## 32367 18

# Something weird here... I think only 1 child is marked as having died in the dataset, but the main paper said at least 45 children died  
 summary(tanzchild$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 3 3 3 3 3 3 32367

hist(tanzchild$AGEDTH)



# Check if age of death is recorded multiple times for each child who died  
 tanzchild %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 1 x 2  
## SUBJID N  
## <int> <int>  
## 1 60212 18

# Recorded multiple times per subject but only for SUBJID 60212

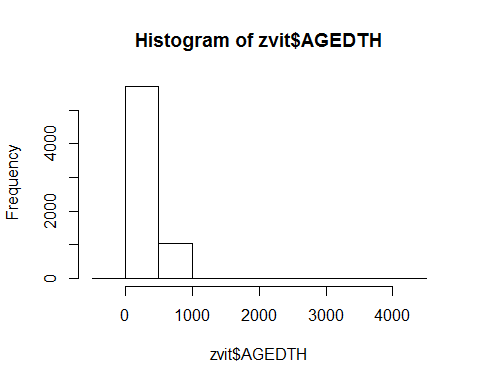
# ZVITAMBO  
 zvit$AGEDTH <- as.numeric(zvit$AGEDTH)  
 table(!is.na(zvit$AGEDTH)) #6778 deaths

##   
## FALSE TRUE   
## 100763 6778

summary(zvit$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -8.0 100.0 222.0 276.1 408.0 4161.0 100763

hist(zvit$AGEDTH)



# Also measured cause of death, CAUSEDTH  
   
 # Check if age of death is recorded multiple times for each child who died  
 zvit %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 1,067 x 2  
## SUBJID N  
## <int> <int>  
## 1 100101 4  
## 2 100241 10  
## 3 100671 16  
## 4 100751 1  
## 5 100781 2  
## 6 100791 9  
## 7 100881 3  
## 8 100981 8  
## 9 101011 1  
## 10 101051 14  
## 11 101481 6  
## 12 101571 20  
## 13 101741 14  
## 14 101851 10  
## 15 101921 11  
## 16 101941 13  
## 17 102041 14  
## 18 102121 12  
## 19 102271 7  
## 20 102331 2  
## 21 102601 3  
## 22 102861 7  
## 23 103141 18  
## 24 103231 4  
## 25 103511 7  
## 26 103601 6  
## 27 103701 8  
## 28 103771 4  
## 29 103831 6  
## 30 104191 14  
## 31 104281 5  
## 32 104381 7  
## 33 104431 4  
## 34 104461 4  
## 35 104561 13  
## 36 104651 5  
## 37 104711 22  
## 38 104761 9  
## 39 104851 5  
## 40 105151 1  
## # ... with 1,027 more rows

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(zvit$deathErrorFlag[!is.na(zvit$AGEDTH)]) #207

##   
## FALSE TRUE   
## 6571 207

#ISSUE - 207 children have measurements after their recorded age of death  
   
 # subset data to children with death error flags  
 df\_error <- zvit %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <int> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 101571 T 248 639 639 T   
## 2 102121 T 202 378 378 T   
## 3 103511 T 557 663 663 T   
## 4 104461 T 2.00 641 641 T   
## 5 105541 T 98.0 180 180 T   
## 6 105591 T 169 275 275 T   
## 7 107881 T 8.00 222 222 T   
## 8 108011 T 89.0 457 457 T   
## 9 108901 T 362 548 548 T   
## 10 109011 T 86.0 548 548 T   
## 11 109301 T 222 427 427 T   
## 12 109481 T 40.0 548 548 T   
## 13 109571 T 545 554 554 T   
## 14 110001 T 91.0 106 106 T   
## 15 110111 T 6.00 180 180 T   
## 16 112321 T 67.0 91 91.0 T   
## 17 112741 T 134 639 639 T   
## 18 113211 T 659 727 727 T   
## 19 113281 T 466 1054 1054 T   
## 20 113761 T 14.0 180 180 T

#make indicator for any death  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- zvit %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died))  
   
 head(df\_death, 20)

## # A tibble: 20 x 5  
## # Groups: SUBJID [3]  
## SUBJID AGEDAYS WHZ AGEDTH died   
## <int> <int> <dbl> <dbl> <lgl>  
## 1 100101 1 - 0.470 74.0 T   
## 2 100101 3 NA 74.0 T   
## 3 100101 43 0.160 74.0 T   
## 4 100101 61 NA 74.0 T   
## 5 100241 1 0.100 514 T   
## 6 100241 61 2.82 514 T   
## 7 100241 107 NA 514 T   
## 8 100241 145 3.64 514 T   
## 9 100241 168 1.97 514 T   
## 10 100241 307 NA 514 T   
## 11 100241 347 0.210 514 T   
## 12 100241 495 NA 514 T   
## 13 100241 496 NA 514 T   
## 14 100241 497 NA 514 T   
## 15 100671 1 - 3.85 483 T   
## 16 100671 26 NA 483 T   
## 17 100671 111 3.64 483 T   
## 18 100671 189 1.79 483 T   
## 19 100671 367 NA 483 T   
## 20 100671 368 NA 483 T

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

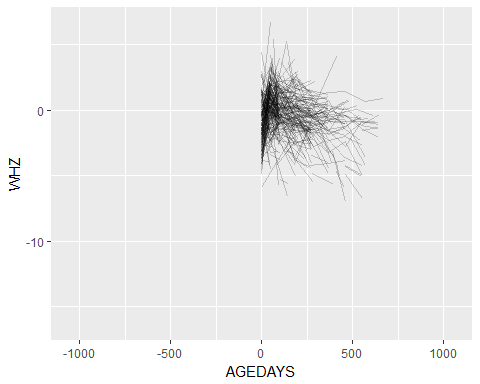
## [1] -0.47 NA 0.16 NA 0.10 2.82 NA 3.64 1.97 NA 0.21  
## [12] NA NA NA -3.85 NA 3.64 1.79 NA NA NA NA  
## [23] NA NA NA NA NA NA -7.32 NA -2.69 -0.30 NA  
## [34] 0.84 0.48 0.90 NA NA -0.58 NA NA NA -9.11 NA  
## [45] NA -2.76 NA NA NA NA NA NA 0.19 -1.85 NA  
## [56] NA NA 1.46 NA NA NA NA 1.58 NA 0.34 NA  
## [67] NA NA -0.80 -1.37 -0.83 NA NA NA -0.48 NA NA  
## [78] 0.88 NA -1.16 NA NA NA NA NA NA NA NA  
## [89] NA NA NA NA NA -4.87 2.13 NA 0.18 0.34 NA  
## [100] NA NA -0.85 NA NA NA -4.41 NA -0.06 NA NA  
## [111] -1.28 NA -0.81 NA NA 0.24 NA 0.52 NA NA NA  
## [122] NA NA NA NA -0.77 NA NA 1.22 NA 1.57 -1.05  
## [133] NA NA NA -1.38 0.15 -1.48 -1.22 -0.46 NA 2.77 NA  
## [144] NA NA NA NA 0.22 NA 2.00 NA -1.40 NA NA  
## [155] NA -0.81 NA NA 1.20 NA -1.82 -2.24 NA NA NA  
## [166] NA -0.51 NA NA -1.40 -2.31 -2.50 NA 1.22 NA 0.88  
## [177] NA NA -0.44 NA NA NA NA NA NA 1.91 NA  
## [188] NA NA 1.03 NA -0.06 NA -0.11 -0.31 0.06 -0.36 NA  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 3879 missing WHZ

##   
## FALSE TRUE   
## 3879 2692

#Plot WHZ trajectories before death among those who died and w/o death errors  
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

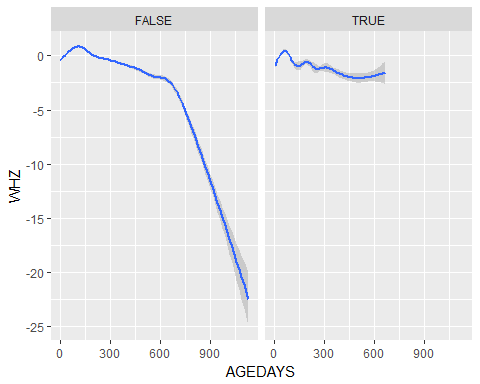
## Warning: Removed 2032 rows containing missing values (geom\_path).



# warning Removed 2032 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(zvit) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 43517 rows containing non-finite values (stat\_smooth).



# warning Removed 43517 rows containing non-finite values

# PROVIDE  
 table(provide$DEAD) # 404 deaths

##   
## 0 1   
## 155865 404

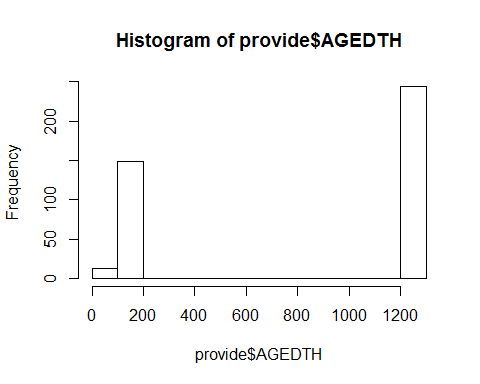
provide$AGEDTH <- as.numeric(provide$AGEDTH)  
 table(!is.na(provide$AGEDTH)) # 404

##   
## FALSE TRUE   
## 155865 404

summary(provide$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 11.0 194.0 1216.0 804.6 1216.0 1216.0 155865

hist(provide$AGEDTH)



# Check if age of death is recorded multiple times for each child who died  
 provide %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 5 x 2  
## SUBJID N  
## <int> <int>  
## 1 1317 244  
## 2 2757 86  
## 3 5871 3  
## 4 8800 62  
## 5 9707 9

# Recorded multiple times per subject  
   
 # Create variable with maximum age per child  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
 # create indicator for last observation for a child using age   
 provide <- provide %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
 # Check that the age of death is later than the oldest measurement on each child  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(provide$deathErrorFlag[!is.na(provide$AGEDTH)]) # 0! :D

##   
## FALSE   
## 404

#No death error flags!  
   
 # subset data to children with death error flags  
 df\_error <- provide %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 0 x 6  
## # Groups: SUBJID [0]  
## # ... with 6 variables: SUBJID <int>, last <lgl>, AGEDTH <dbl>,  
## # AGEDAYS <int>, maxAge <dbl>, deathErrorFlag <lgl>

#make indicator for any death  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- provide %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died))  
   
 head(df\_death, 20)

## # A tibble: 20 x 5  
## # Groups: SUBJID [1]  
## SUBJID AGEDAYS WHZ AGEDTH died   
## <int> <int> <dbl> <dbl> <lgl>  
## 1 1317 7 - 1.66 1216 T   
## 2 1317 8 NA 1216 T   
## 3 1317 12 NA 1216 T   
## 4 1317 15 NA 1216 T   
## 5 1317 16 NA 1216 T   
## 6 1317 17 NA 1216 T   
## 7 1317 19 NA 1216 T   
## 8 1317 22 NA 1216 T   
## 9 1317 26 NA 1216 T   
## 10 1317 27 NA 1216 T   
## 11 1317 29 NA 1216 T   
## 12 1317 32 NA 1216 T   
## 13 1317 33 NA 1216 T   
## 14 1317 36 NA 1216 T   
## 15 1317 51 1.09 1216 T   
## 16 1317 54 NA 1216 T   
## 17 1317 57 NA 1216 T   
## 18 1317 61 NA 1216 T   
## 19 1317 62 NA 1216 T   
## 20 1317 64 NA 1216 T

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

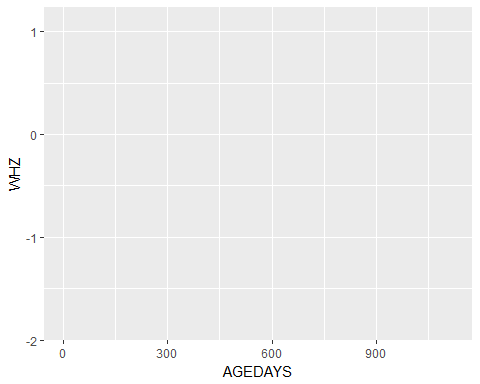
## [1] -1.66 NA NA NA NA NA NA NA NA NA NA  
## [12] NA NA NA 1.09 NA NA NA NA NA NA NA  
## [23] NA NA NA NA NA NA NA 1.05 NA NA NA  
## [34] NA NA 0.84 NA NA NA NA -0.32 NA NA NA  
## [45] NA NA NA NA NA NA -0.52 NA NA NA -0.87  
## [56] NA NA NA NA NA NA NA NA NA NA NA  
## [67] NA NA NA NA NA NA -1.13 NA NA NA NA  
## [78] NA NA NA NA NA NA NA NA NA NA NA  
## [89] NA NA NA NA NA NA NA NA NA NA NA  
## [100] NA NA NA NA NA NA NA NA NA NA NA  
## [111] NA NA -0.94 NA NA -1.52 NA NA NA NA NA  
## [122] NA NA NA NA NA NA NA NA NA NA NA  
## [133] NA NA NA NA -0.80 NA NA -1.34 NA NA NA  
## [144] NA NA NA NA NA NA NA NA NA -0.61 NA  
## [155] NA NA NA NA NA NA NA NA NA NA NA  
## [166] NA NA NA NA NA NA NA NA NA NA NA  
## [177] NA NA NA -1.42 NA NA NA NA NA NA NA  
## [188] NA NA NA NA NA NA NA NA NA NA NA  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 369 missing WHZ

##   
## FALSE TRUE   
## 369 35

#Plot WHZ trajectories before death among those who died and w/o death errors  
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

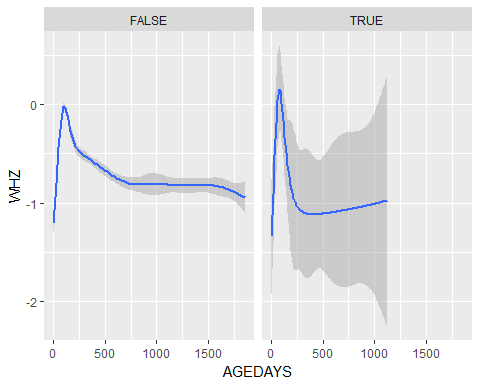
## Warning: Removed 20 rows containing missing values (geom\_path).



# warning removed 20 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(provide) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) +  
 facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 144666 rows containing non-finite values (stat\_smooth).



# warning Removed 144666 rows containing non-finite values